

University of Groningen

Feasibility, test-retest reliability, and interrater reliability of the Modified Ashworth Scale and Modified Tardieu Scale in persons with profound intellectual and multiple disabilities

Waninge, A.; Rook, R.A.; Dijkhuizen, A.; Gielen, E.; van der Schans, C.P.

Published in:
Research in Developmental Disabilities

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

Publication date:
2011

[Link to publication in University of Groningen/UMCG research database](#)

Citation for published version (APA):

Waninge, A., Rook, R. A., Dijkhuizen, A., Gielen, E., & van der Schans, C. P. (2011). Feasibility, test-retest reliability, and interrater reliability of the Modified Ashworth Scale and Modified Tardieu Scale in persons with profound intellectual and multiple disabilities. *Research in Developmental Disabilities*, 32(2), 613-620.

Copyright

Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

Take-down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): <http://www.rug.nl/research/portal>. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.



Feasibility, test–retest reliability, and interrater reliability of the Modified Ashworth Scale and Modified Tardieu Scale in persons with profound intellectual and multiple disabilities

A. Waninge^{a,b,*}, R.A. Rook^{b,1}, A. Dijkhuizen^{b,1}, E. Gielen^{c,2}, C.P. van der Schans^{b,1}

^a Royal Dutch Visio De Brink, Veenweg 20, 9481 TJ, Vries, The Netherlands

^b Hanze University Groningen, Applied Sciences in Health Care and Nursing, Eyseniusplein 18, 9714 CE Groningen, The Netherlands

^c ASVZ Sliedrecht, Postbus 121, 3360 AC Sliedrecht, The Netherlands

ARTICLE INFO

Article history:

Received 31 October 2010

Accepted 13 December 2010

Available online 12 January 2011

Keywords:

Muscle tone

Spasticity

Modified Ashworth Scale

Modified Tardieu Scale

Persons with PIMD

Profound intellectual and multiple disabilities

ABSTRACT

Caregivers of persons with profound intellectual and multiple disabilities (PIMD) often describe the quality of the daily movements of these persons in terms of flexibility or stiffness. Objective outcome measures for flexibility and stiffness are muscle tone or level of spasticity. Two instruments used to grade muscle tone and spasticity are the Modified Ashworth Scale (MAS) and the Modified Tardieu Scale (MTS). To date, however, no research has been performed to determine the psychometric properties of the MAS and MTS in persons with PIMD. Therefore, the purpose of this study was to determine the feasibility, test–retest reliability, and interrater reliability of the MAS and MTS in persons with PIMD. We assessed 35 participants on the MAS and MTS twice, first for the test and second a week later for the retest. Two observers performed the measurements. Feasibility was assessed based on the percentage of successful measurements. Test–retest and interrater reliability were determined by using the Wilcoxon signed rank test, intraclass correlation coefficients (ICC), Spearman's correlation, and either limits of agreement (LOA) or quadratically weighted kappa. The feasibility of the measurements was good, because an acceptable percentage of successful measurements were performed. MAS measurements had substantial to almost perfect quadratically weighted kappa (>0.8) and an acceptable ICC (>0.8) for both inter- and intrarater reliability. However, MTS measurements had insufficient ICCs, Spearman's correlations, and LOAs for both inter- and interrater reliability. Our data indicated that the feasibility of the MAS and MTS for measuring muscle tone in persons with PIMD was good. The MAS had sufficient test–retest and interrater reliability; however, the MTS had an insufficient test–retest and interrater reliability in persons with PIMD. Thus, the MAS may be a good method for evaluating the quality of daily movements in persons with PIMD. Providing test administrators with training and clear instructions will improve test reliability.

© 2010 Elsevier Ltd. All rights reserved.

* Corresponding author at: Royal Dutch Visio De Brink, Veenweg 20, 9481 TJ, Vries, The Netherlands. Tel.: +31 (0)592 393787; fax: +31 (0)592 39 38 89.

E-mail addresses: alywaninge@visio.org (A. Waninge), r.a.rook@st.hanze.nl (R.A. Rook), a.dijkhuizen@pl.hanze.nl (A. Dijkhuizen), EGielen@asvz.nl (E. Gielen), c.p.van.der.schans@pl.hanze.nl (C.P. van der Schans).

¹ Tel.: +31 (0)50 5953685; fax: +31 (0)50 595 56 78.

² Tel.: +31 (0)184 49 1200; fax: +31 (0)184 49 1500.

1. Introduction

Persons with profound intellectual and multiple disabilities (PIMD) generally have very limited mobility, always use a wheelchair (Van der Putten, Vlaskamp, Reynders, & Nakken, 2005), and often have a Gross Motor Function Classification System (GMFCS) level of IV or V (Palisano et al., 2000). Spasticity, dyskinesia, or ataxia with hypotony frequently occurs in individuals in these GMFCS levels (Shevell, Dagenais, & Hall, 2009). Lance (1990) defined spasticity as “a motor disorder, characterised by a velocity-dependent increase in tonic stretch reflexes (muscle tone) with exaggerated tendon jerks, resulting from hyper-excitability of the stretch reflex as one component of the upper motor neurone syndrome”. Muscle tonus or muscle tone is “the state of activity or tension of a muscle beyond that related to its physical properties, that is, its active resistance to stretch. In skeletal muscle, tonus is dependent upon efferent innervation” (Stedman, 1999).

Persons with PIMD are at risk for a variety of limitations in daily functioning (Evenhuis, Sjoukes, Koot, & Kooijman, 2009), such as inactivity, unsteady movement, and diminished initiative. However, research into the quality of daily movements of persons with PIMD is limited, and related knowledge is scarce. Caregivers of persons with PIMD often describe the quality of daily movements in terms of flexibility or stiffness. Objective outcome measures for flexibility and stiffness are muscle tone or level of spasticity.

Bohannon and Smith (1987) introduced the Modified Ashworth Scale (MAS) as a scale for grading spasticity. The MAS is a clinical measure of muscle tone and a nominal-level measure of resistance to passive movement (Pandyan et al., 1999). The reliability of the scale appears to be better for measuring muscle tone of upper limbs (Pandyan et al., 1999). Although one study found the reliability of the MAS to be very good (kappa was 0.84 for interrater and 0.83 for intrarater comparisons) (Gregson, Leathley, Moore, Sharma, Smith, & Watkins, 1999), other studies found it to be insufficient (Ansari, Naghdi, Arab, & Jalaie, 2008; Clopton et al., 2005; Mutlu, Livanelioglu, & Gunel, 2008; Yam & Leung, 2006).

Haugh, Pandayan, and Johnson (2006) suggested that the Modified Tardieu Scale (MTS) is a more appropriate clinical measure of spasticity than the MAS. The MTS assesses resistance to passive movement at both slow and fast speeds, and therefore adheres more closely to Lance's definition of spasticity (Haugh et al., 2006). Both parameters of the MTS have excellent intrarater and interrater reliability in children with cerebral palsy (Gracies et al., 2010). However, as with the MAS, other studies found the MTS to have insufficient reliability (Ansari, Naghdi, Hasson, Azarsa, & Azarnia, 2008; Mackey, Walt, Lobb, & Stott, 2004; Yam & Leung, 2006).

Both the MTS and the MAS show sufficient interrater and intrarater reliability in adults with intellectual disabilities (Gielen, 2005), but the MTS seems to be more feasible and reliable than the MAS. The MTS also shows more reliability than the MAS in adults with severe brain injury and spasticity (Mehrholtz et al., 2005). Haugh et al. (2006) stated that further studies need to be undertaken to clarify the validity and reliability of the MTS and the MAS for a variety of muscle groups in adult neurological patients. Thus far, no research has been performed to determine the psychometric properties of the MTS and MAS in persons with PIMD. Therefore, the purpose of this study was to determine the feasibility, test–retest reliability, and interrater reliability of the MAS and MTS in persons with PIMD.

2. Methods

2.1. Participants

We asked the representatives of 42 persons with PIMD for written permission for these persons to participate in our study. Forty representatives gave permission. After informed consent was obtained, the subjects were screened based on an examination by both a special needs physician and a behavioral scholar. The screening exclusion criteria were severe psychological problems or somatic diseases, which were defined as chronic diseases and/or diseases that do not resolve in the short term. Two persons were excluded because they exhibited one of these problems or diseases. The exclusion criteria at the time the measurements were being performed were general illness or fever; taking antibiotics; recently started taking muscle relaxants; worsening of asthma, epilepsy (recent insult or epileptic fits); fresh wound(s)/bruise(s) or other factors causing pain during movement; or stress due to the subject's behavior just before the measurement date. Three persons were excluded because they exhibited one of these criteria. Fig. 1 presents the sampling scheme of persons included in the study.

The participants with PIMD were classified as GMFCS IV or V (Palisano et al., 2000). Furthermore, the intellectual level or intelligence quotient (IQ) of each participant was classified according to the International Classification of Diseases (ICD-10) of the World Health Organization (WHO, 1992). The presence or absence of epilepsy was also recorded, because we assumed that seizures greatly affect muscle tone. We also classified the visual impairments of the participants according to WHO guidelines (WHO, 2001). Finally, the presence or absence of orthopedic disorders was recorded.

2.2. Ethical statement

The study was performed in agreement with the guidelines of the Helsinki Declaration as revised in 1975. Permission to carry out the study was obtained from the institutional ethics committee. Informed consent was obtained from legal representatives of the participants, because all participants were unable to give consent. The measurements were performed in accordance with the guidelines of the Dutch Society for Doctors in the Care for people with an Intellectual Disability (NVAZ), which are outlined in a code called “Resistance among people with an intellectual disability in the

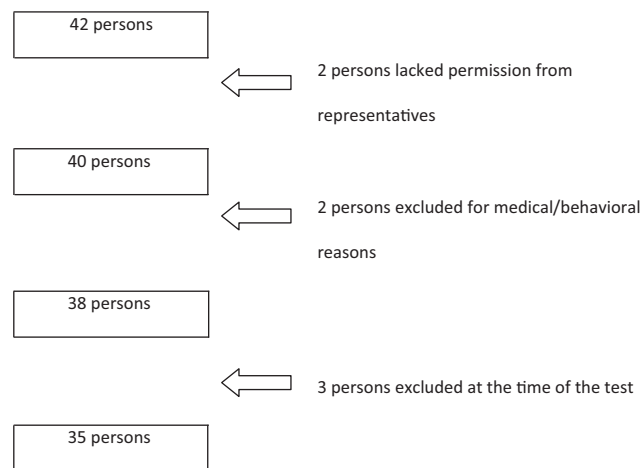


Fig. 1. Sampling scheme of subjects included in the study.

framework of the Act Governing Medical-Scientific Research Involving Humans" (NVAZ, 1999). The purpose of this code is to guide doctors in assessing resistance in persons with an intellectual disability. In line with this code, a participant's consistent distress or unhappiness was interpreted as a sign of lack of assent, and further participation in the study was reconsidered.

2.3. Design

The muscle tone and spasticity of 35 participants were measured twice (test and retest) with the MAS and MTS. The retest was conducted one week after the initial test. The participants were first assessed with the MAS and afterwards with the MTS. For each participant, both measurements were conducted at the same time of day and under the same conditions. Two observers performed the measurements. The interrater reliability of the MAS and MTS was determined from the measurements of the two observers. The test–retest reliability of the MAS and MTS was determined by using the test–retest measurements of observer 1.

2.4. Measures

Prior to the measurements, the observers and personal guides of the participants completed a checklist containing the exclusion criteria. Both observers were present at the time of the measurements. So that testing would not cause additional stress to the participants, we made sure that the participants were familiar with observer 2. We created a protocol describing how to administer the MAS and MTS based on the protocol of Gielen (2005). Gracies et al. (2010) stated that training was associated with a highly significant improvement in reliability, so the observers were trained on how to perform the protocol. The training consisted of a brief explanation of the protocol and practical exercises. During the practical exercises, the results were compared and discussed.

2.4.1. Modified Ashworth Scale (MAS)

The MAS was carried out as follows. During five repetitions of a passive motion within one second, resistance was scored on the following 6-point scale (Bohannon & Smith, 1987):

0 = No increased resistance; 1 = Slightly increased resistance (catch followed by relaxation or minimal resistance at the end of the range of motion); 1+ = Slightly increased resistance (catch followed by minimal resistance throughout less than half of the range of motion); 2 = Clear resistance throughout most of the range of motion; 3 = Strong resistance; passive movement is difficult; 4 = Rigid flexion or extension

Catch is the phenomenon that suddenly a strong resistance occurs during a fast passive movement.

2.4.2. Modified Tardieu Scale (MTS)

The MTS consists of two measurements: R2 and R1 (Mackey et al., 2004). In the present study, we measured the most restricted joint motion of both the elbow and the knee. A goniometer was used for measuring the range of motion. The measurements were accurate to the 5-degree level. The R2 measurement consisted of slow motion performed within one second. The range of motion was measured with a goniometer. The R1 measurement consisted of fast motion performed within half a second. The range of motion immediately after the catch was measured with a goniometer.

2.5. Data analyses

The data were analyzed using SPSS 15.0. The distribution of the data was determined and checked for normal distribution.

2.5.1. Feasibility

To assess feasibility, we compared the number of successful measurements per task to the total number of measurements. Since it only makes sense to use a test if a reasonable percentage of successful measurements can be made, this aspect of feasibility was considered to be sufficient if 85% of the measurements were successful (Lemmink, 1996; Malmberg et al., 2002).

2.5.2. Test–retest reliability

Firstly, to determine whether significant differences between test and retest measurements exist, we analyzed the differences using the *t*-test or, in case of non-normally distributed data, the Wilcoxon signed rank test. The level of statistical significance was set at 0.05.

Secondly, intraclass correlation coefficients (ICC; two-way random, absolute agreement) of measurements 1 and 2 were computed. Reliability was considered to be acceptable if the ICC was greater than 0.75 and the 95% confidence interval (CI) was 0.3 or less. Reliability was considered to be very good if the ICC was greater than 0.9 (Portney & Watkins, 2000).

Thirdly, Spearman correlation coefficients of measurements 1 and 2 were computed. Spearman's correlation was used because the data were not normally distributed. A correlation of 0.61 or more is considered good (Feinstein, 1987).

Fourthly, for the MTS, limits of agreement (LOA) between measurements 1 and 2 were calculated according to the procedure described by Bland and Altman (1986). LOAs were expressed together with the mean differences between measurements 1 and 2, and were judged whether they were narrow enough for the test to be of practical use, according to Atkinson and Nevill (1998). For the MAS, quadratically weighted kappa for measurements 1 and 2 was calculated. The quadratically weighted kappa is a measure of the proportion of agreement greater than that expected by chance. Values of kappa below 0.40 are generally considered to be clinically unacceptable, those within 0.41–0.60 to be moderate, those within 0.61–0.80 to be substantial, and those 0.81–1.00 to be almost perfect (Sim & Wright, 2005). To obtain 95% (CIs) for the weighted kappa coefficients, we used the adjusted bootstrap percentile (BCa) method (Davison & Hinkley, 1997; Efron, 1987) by employing the statistical programming language R (R Development Core Team, 2009).

Finally, the test–retest reliability of the MTS was considered reliable if (1) there were no significant differences between the test and retest measurements; (2) ICC was acceptable, as described above; (3) Spearman correlation coefficient was acceptable, as described above; and (4) LOA was acceptable, as described above. The test–retest reliability of the MAS was considered reliable if (1) there were no significant differences between the test and retest measurements; (2) ICC was acceptable, as described above; (3) Spearman correlation coefficient was acceptable, as described above; and (4) quadratically weighted kappa was almost perfect, with a 95% CI from substantial to almost perfect, as described above.

2.5.3. Interrater reliability

Firstly, to determine whether significant differences between the measurements of observers 1 and 2 exist, we analyzed the differences across measurements using a *t*-test or, in case of non-normally distributed data, the Wilcoxon signed rank test. The level of statistical significance was set at 0.05.

Secondly, the ICCs (two-way random, absolute agreement) of the measurements of observers 1 and 2 were computed. Reliability was considered to be acceptable if the ICC was greater than 0.75. Reliability was considered to be very good if the ICC was greater than 0.9 (Portney & Watkins, 2000).

Thirdly, Spearman correlation coefficients of the measurements of observers 1 and 2 were computed. Spearman's correlation was used because the data were not normally distributed. A correlation of 0.61 or more is considered good (Feinstein, 1987).

Fourthly, for the MTS, LOAs between the measurements of observer 1 and 2 were calculated according to the procedure described by Bland and Altman (1986). LOAs were expressed together with the mean differences between the measurements of observers 1 and 2, and were judged whether they were narrow enough for the test to be of practical use, according to Atkinson and Nevill (1998). For the MAS, quadratically weighted kappa for the measurements of observers 1 and 2 was calculated. Values of kappa below 0.40 are generally considered to be clinically unacceptable, those within 0.41–0.60 to be moderate, those within 0.61–0.80 to be substantial, and those within 0.81–1.00 to be almost perfect (Sim & Wright, 2005). We calculated 95% (CIs) for the weighted kappa coefficients (adjusted BCa method) (Davison & Hinkley, 1997; Efron, 1987) by employing the statistical programming language R (R Development Core Team, 2009).

Finally, the interrater reliability of the MTS was considered acceptable if (1) there were no significant differences between the measurements of observers 1 and 2; (2) LOA was acceptable, as described above; (3) Spearman correlation coefficient was acceptable as described above; and (4) ICC was acceptable, as described above. The interrater reliability of the MAS was considered reliable if (1) there were no significant differences between the measurements of observers 1 and 2; (2) ICC was acceptable, as described above; (3) Spearman correlation coefficient was acceptable, as described above; and (4) quadratically weighted kappa was almost perfect, with a 95% CI from substantial to almost perfect, as described above.

Table 1
Characteristics of the study population.

Characteristics											
Intellectual disability		Visual impairment		GMFCS level		Spasticity		Orthopedic defects		Epilepsy	
Severe	22	Blind/severe	25	Level 4	11	Yes	17	Yes	30	Yes	32
Profound	13	Partially	10	Level 5	24	No	18	No	5	No	3

3. Results

The data were not normally distributed; therefore, non-parametric tests were used to analyze the data. In all, 35 subjects participated in this study; 22 were male (62.9%), and 13 were female (37.1%). The mean age (SD) of the men was 35 (15) years, and that of the women was 31 (12) years. The characteristics of the study population are shown in Table 1.

3.1. Feasibility

3.1.1. Percentage of successful measurements

The percentages of successful measurements are shown in Table 2. Both the MAS and MTS showed a sufficient percentage of successful measurements.

3.2. Test–retest reliability of the MTS

Table 3 summarizes the statistical analyses for measurements 1 and 2 of the MTS.

There were no significant differences between measurements 1 and 2 ($p < 0.05$). The ICC showed acceptable agreement for the R2 measurement of the arm and the R1 measurement of the leg. The ICC for the R1 measurement of the arm and the R2 measurement of the leg was not acceptable. The LOAs for both arm and leg measurements compared to median values were considerably large, which is clinically unacceptable. The Spearman correlation coefficient showed good correlation between the test–retest measurements for both arm and leg, except for the R2 measurement of the leg which was not acceptable.

3.3. Test–retest reliability of the MAS

Table 4 summarizes the statistical analyses for measurements 1 and 2 of the MAS.

There were no significant differences between measurements 1 and 2 ($p < 0.05$). The ICC showed acceptable agreement between MAS measurements 1 and 2 for the arm and leg. The quadratically weighted kappa was almost perfect for the arm measurements and substantial for the leg measurements; the 95% CI values were substantial to

Table 2
Percentages of successful measurements for the MTS and MAS.

	Successful measurements week 1 Observer 1	Successful measurements week 2 Observer 1
Modified Tardieu Scale (MTS)		
Arm R2	100%	97.1%
Arm R1	100%	97.1%
Leg R2	97.1%	94.3%
Leg R1	97.1%	94.3%
Modified Ashworth Scale (MAS)		
Arm	100%	97.1%
Leg	97.1%	94.3%

Table 3
Summary of the statistical analyses for measurements 1 and 2 of the MTS for test–retest reliability.

Modified Tardieu Scale (MTS)	Number of subjects	Median M1 (min–max)	Median M2 (min–max)	P level Wilcoxon	ICC	Mean difference \pm LOA	Spearman coefficient
Arm R2	34	30 (0–95)	32.50 (0–95)	0.592	0.815	2.353 \pm 35.2	0.792
Arm R1	34	55 (0–100)	57.50 (0–90)	0.890	0.627	6.029 \pm 57.7	0.624
Leg R2	33	70 (0–135)	70 (0–120)	0.779	0.741	1.818 \pm 44	0.402
Leg R1	33	67.50 (0–110)	75 (0–115)	0.089	0.850	–5.75 \pm 29.8	0.680

Table 4

Summary of the statistical analyses for measurements 1 and 2 of the MAS for test–retest reliability.

Modified Ashworth Scale (MAS)	Number of subjects	Median M1 (min–max)	Median M2 (min–max)	P level Wilcoxon	ICC	Quadratically weighted kappa	Spearman coefficient
Arm	35	1 0–4	1 0–4	0.430	0.853	0.82 0.63–0.92	0.808
Leg	33	1+ 0–4	1 0–4	0.771	0.813	0.78 0.60–0.89	0.757

Table 5

Summary of the statistical analyses for the MTS measurements of observers 1 and 2.

Modified Tardieu Scale (MTS)	Number of subjects	Median O1 (min–max)	Median O2 (min–max)	P level Wilcoxon	ICC	Mean difference \pm LOA	Spearman coefficient
Arm R2	24	30 (0–95)	30 (0–110)	0.886	0.806	0.00 \pm 40	0.813
Arm R1	24	55 (0–100)	70 (0–100)	0.540	0.851	5.870 \pm 38	0.825
Leg R2	23	70 (0–135)	60 (0–115)	0.730	0.766	–0.91 \pm 54	0.696
Leg R1	23	67.50 (0–110)	60 (0–120)	0.924	0.877	–0.45 \pm 60	0.726

Table 6

Summary of the statistical analyses for the MAS measurements of observers 1 and 2.

Modified Ashworth Scale (MAS)	Number of subjects	Median O1 (min–max)	Median O2 (min–max)	P level Wilcoxon	ICC	Quadratically weighted kappa	Spearman coefficient
Arm	23	1 0–4	1 0–4	0.350	0.894	0.88 0.75–0.97	0.907
Leg	23	1+ 0–4	1+ 0–4	0.390	0.895	0.88 0.67–0.96	0.858

almost perfect. The Spearman correlation coefficient demonstrated clear agreement between measurements 1 and 2 for the arm and leg.

3.4. Interrater reliability of the MTS

Table 5 summarizes the statistical analyses for the MTS measurements of observers 1 and 2. There were no significant differences between the measurements made by observers 1 and 2 ($p < 0.05$). The ICC showed acceptable agreement between all the R2 and R1 measurements of observer 1 and 2. The LOAs of both arm and leg measurements compared to median values were considerably large, which is clinically not acceptable. The Spearman correlation coefficients between the R2 and R1 measurements of observer 1 and 2 were acceptable.

3.5. Interrater reliability of the MAS

Table 6 summarizes the statistical analyses for MAS measurements of observers 1 and 2. There were no significant differences between the measurements taken by observers 1 and 2 ($p < 0.05$). The ICC shows an acceptable agreement between the measurements of the arm and the leg of observers 1 and 2. The quadratically weighted kappa is almost perfect for the measurements of observers 1 and 2 of both the arm and the leg, the 95% CI are substantial to almost perfect. The Spearman correlation coefficient between the measurements of the arm and the leg of observers 1 and 2 is acceptable.

4. Discussion

The purpose of our study was to determine the feasibility, the test–retest reliability, and interrater reliability of the MAS and MTS in persons with PIMD. Our results demonstrated that the feasibility of the measurements was good, as an acceptable percentage of successful measurements were performed. The interrater reliability of the MAS was sufficient, with a substantial to almost perfect quadratically weighted kappa and an acceptable ICC. However, we found the reliability of the MTS not to be clinically acceptable. Although the ICC indicated that the interrater reliability of the MTS was sufficient, the LOAs for both arm and leg measurements relative to median values were considerably large, which is clinically not acceptable.

The MAS showed a sufficient test–retest reliability, with a substantial to almost perfect quadratically weighted kappa and an acceptable ICC. However, the test–retest reliability of the MTS was not sufficient due to its insufficient ICC, Spearman's correlations, and clinically unacceptable LOAs for both arm and leg measurements.

In our target group, the MAS showed a better test–retest and interrater reliability than the MTS, which contradicts the results of Gielen (2005) and Mehrholz et al. (2005). In the study of Gielen (2005), for the MAS the test–retest reliability

calculated with Spearman's rho ranged from 0.66 to 0.81 (our study: 0.76–0.86) and the interrater reliability from 0.67 to 0.80 (our study: 0.86–0.91). For the MTS, Gielen's Spearman's rho was slightly better, with a range of 0.70–0.88 for intrarater reliability (our study: 0.40–0.79) and 0.70–0.82 for interrater reliability (our study: 0.70–0.83). As mentioned in the introduction, these contradictory findings also occurred in other target groups. All participants in our study population had impaired vision. Impaired vision may have contributed to differences in the faster movements of the MTS R2 measurements. Our subjects could not anticipate fast movements as well as their peers without visual impairments.

Compared to the study of Clopton et al. (2005) in children with hypertonia, our ICC values for the MAS (0.81–0.85 for intrarater; 0.89–0.89 for interrater) were more sufficient than their ICC values (0.5–0.75 for intrarater; <0.5 for interrater). Mutlu et al. (2008) performed a reliability study of the MAS in children with spastic cerebral palsy and found an intrarater reliability of 0.36–0.83 (ICC) and an interrater reliability of 0.54–0.78 (ICC), which is also less reliable than our ICC values for the MAS.

Our quadratically weighted kappa score for the intrarater reliability of the MAS was higher (interrater reliability 0.88; intrarater reliability 0.78–0.82) than the kappa (interrater reliability 0.51; intrarater reliability 0.59) of Ansari, Naghdi, Arab, et al. (2008), which tested the MAS in persons with hemiplegia. In the study of Gregson et al. (1999), which involved post-stroke patients, the kappa score for interrater reliability was 0.84 and for intrarater reliability was 0.83, which is comparable to the quadratically weighted kappa scores of the present study.

In our target group, the MTS had ICCs of 0.76–0.88 for interrater reliability and 0.63–0.85 for intrarater reliability, which is better than the interrater scores of Ansari, Naghdi, Hasson, et al. (2008) in patients with hemiplegia (<0.56). In general, in the present study the interrater reliability of the MTS was better than the intrarater reliability. This difference may be due to changes in the condition of the participants at the time of the two measurements. This premise is also supported by the relatively large LOA outcomes, which indicated that there was too much variation at the individual level. The LOAs for both arm and leg measurements compared to median values was considerably large; most of the measurements departed more than 50% from the median. The LOAs were not narrow enough to indicate agreement between the two measurements of an individual. Therefore, we concluded that the LOAs for the MTS measurements were clinically unacceptable. Haugh et al. (2006) suggested that the MTS is a more appropriate clinical measure of spasticity than the MAS. Therefore, we determined whether the LOA for MTS measurements of persons with spasticity was less than the LOA for measurements of persons without spasticity. However, the LOAs did not differ between these groups.

The purpose of our study was to determine the feasibility and reliability of the MAS and MTS in persons with PIMD. However, the validity of these instruments for measuring either spasticity or muscle tone was beyond the scope of this study. After the start of our study a manuscript was published entitled, "Stop using the Ashworth Scale for the assessment of spasticity" (Fleuren et al., 2010). Taking their paper into account, we recommend that the validity of the MAS be examined further in future studies. Doing so, however, raises the issue of which specific parameter must be examined—the validity of measuring spasticity or the validity of measuring the quality of daily movements—since the MAS may be more suitable for measuring the latter. Ghotbi et al. (2009) also obtained reliable measurements with the Modified Ashworth Scale, supporting the recommendation that further research should be done on this instrument.

Given the outcomes of the ICC, quadratically weighted kappa, and the Spearman's correlation for the MAS compared to corresponding values obtained by other studies, the interrater and intrarater reliability of the MAS are sufficient in this target group. Furthermore, the usability of the MAS and MTS according to the observers appeared to be good. The protocol developed for the study functioned well for the observers. The observers were able to perform the tests properly after receiving training on the test protocol. The instructions were clear, and other physical therapists were able to follow the test protocol. The instruments also fit well within physical therapy. Moreover, testing was brief and the instruments were inexpensive.

5. Conclusions

Our research showed that the feasibility of using the MAS and MTS for measuring muscle tone in persons with PIMD is good. In our participants, the MAS showed sufficient test–retest and interrater reliability, whereas the MTS showed insufficient test–retest and interrater reliability. Therefore, the MAS may be a good method for evaluating the quality of daily movements in individuals with PIMD. Providing training and clear instructions on administering the MAS will improve reliability. Further research should aim to examine the validity of the MAS.

6. Recommendations

The feasibility of conducting MAS and MTS measurements in persons with PIMD is good. Adjustments in implementing the MAS and MTS are not necessary. Further research involving more participants may provide additional information.

Acknowledgements

This research was financed by Hanze University Groningen and Royal Dutch Visio De Brink. The authors kindly acknowledge and thank the participants for their participation in this study, their representatives for given permission to

this and the physical therapists of Royal Dutch Visio The Brink, for their participation and for accompanying the participants during the tests. Furthermore, the authors kindly acknowledge W. Krijnen, PhD, for helping with statistical analyses.

References

- Ansari, N. N., Naghdi, S., Arab, T. K., & Jalaie, S. (2008). The interrater and intrarater reliability of the Modified Ashworth Scale in the assessment of muscle spasticity: Limb and muscle group effect. *NeuroRehabilitation*, 23, 231–237.
- Ansari, N. N., Naghdi, S., Hasson, S., Azarsa, M. H., & Azarnia, S. (2008). The Modified Tardieu Scale for the measurement of elbow flexor spasticity in adult patients with hemiplegia. *Brain Injury*, 22, 1007–1012.
- Atkinson, G., & Nevill, A. M. (1998). Statistical methods for assessing measurement error (reliability) in variables relevant to sports medicine. *Sports Medicine*, 26, 217–238.
- Bland, J. M., & Altman, D. G. (1986). Statistical methods for assessing agreement between two methods of clinical measurement. *The Lancet*, 1, 307–310.
- Bohannon, R. W., & Smith, M. B. (1987). Interrater Reliability of a Modified Ashworth Scale of Muscle Spasticity. *Physical Therapy*, 67, 206–207.
- Clopton, N., Dutton, J., Featherston, T., Grigsby, A., Mobley, J., & Melvin, J. (2005). Interrater and intrarater reliability of the Modified Ashworth Scale in children with hypertonia. *Pediatric Physical Therapy*, 17, 268–274.
- Davison, A. C., & Hinkley, D. V. (1997). *Bootstrap methods and their application*. Cambridge University Press.
- Dutch Society for Doctors in the Care for people with an Intellectual Disability (NVAZ). (1999). Resistance among people with an intellectual disability in the framework of the Act Governing Medical-Scientific Research Involving Humans, behavioural code for doctors in the assessment of resistance among people with an intellectual disability.
- Efron, B. (1987). Better bootstrap confidence intervals. *Journal of the American Statistical Association*, 82, 171–185.
- Evenhuis, H. M., Sjoukes, L., Koot, H. M., & Kooijman, A. C. (2009). Does visual impairment lead to additional disability in adults with intellectual disabilities? *Journal of Intellectual Disability Research*, 53, 19–28.
- Feinstein, A. R. (1987). *Clinimetrics* (pp. 184–185). New Haven: Yale University Press.
- Fleuren, J. F., Voerman, G. E., Erren-Wolters, C. V., Snoek, G. J., Rietman, J. S., Hermens, H. J., et al. (2010). Stop using the Ashworth Scale for the assessment of spasticity. *Journal of Neurology, Neurosurgery, and Psychiatry*, 81, 46–52.
- Ghotbi, N., Ansari, N. N., Naghdi, S., Hasson, S., Jamshidpour, B., & Amiri, S. (2009). Inter-rater reliability of the Modified Ashworth Scale in assessing lower limb muscle spasticity. *Brain Injury*, 23, 815–819.
- Gielen, E. J. J. M. (2005). Is spasticiteit te meten bij mensen met een verstandelijke beperking? Een betrouwbaarheids- en validiteitsonderzoek van de Modified Ashworth Scale en de Tardieu Schaal bij mensen met een verstandelijke beperking. Scriptie in het kader van de cursus Scholing in Wetenschap III; Amersfoort, Nederlands Paramedisch Instituut.
- Gracies, J. M., Burke, K., Clegg, N. J., Browne, R., Rushing, C., Fehlings, D., et al. (2010). Reliability of the Tardieu Scale for assessing spasticity in children with cerebral palsy. *Archives of Physical Medicine and Rehabilitation*, 91, 421–428.
- Gregson, J. M., Leathley, M., Moore, A. P., Sharma, A. K., Smith, T. L., & Watkins, C. L. (1999). Reliability of the Tone Assessment Scale and the Modified Ashworth Scale as clinical tools for assessing poststroke spasticity. *Archives of Physical Medicine and Rehabilitation*, 80, 1013–1016.
- Haugh, A. B., Pandyan, A. D., & Johnson, G. R. (2006). A systematic review of the Tardieu Scale for the measurement of spasticity. *Disability and Rehabilitation*, 28, 899–907.
- Lance, J. W. (1990). What is spasticity? *The Lancet*, 335, 606.
- Lemmink, K. A. P. M. (1996). *De Groninger Fitheidstest voor Ouderen*. Ontwikkeling van een meetinstrument. RUG.
- Mackey, A. H., Walt, S. E., Lobb, G., & Stott, N. S. (2004). Intraobserver reliability of the Modified Tardieu Scale in the limb of children with Hemiplegia. *Developmental Medicine and Child Neurology*, 46, 267–272.
- Malmberg, J. J., Miilunpalo, S. I., Vuori, I. M., Pasanen, M. E., Oja, P., & Haapanen-Niemi, N. A. (2002). A health-related fitness and functional performance test battery for middle-aged and older adults: Feasibility and health-related content validity. *Archives of Physical Medicine*, 83, 666–677.
- Mehrholz, J., Wagner, K., Meißner, D., Grundmann, K., Zange, C., Koch, R., et al. (2005). Reliability of the Modified Tardieu Scale and the Modified Ashworth Scale in adults patients with severe brain injury: A comparison study. *Clinical Rehabilitation*, 19, 751.
- Mutlu, A., Livanelioglu, A., & Gunel, M. K. (2008). Reliability of Ashworth and Modified Ashworth scales in children with spastic cerebral palsy. *BMC Musculoskeletal Disorders*, 10, 9–44.
- Palisano, R., Hanna, S. E., Rosenbaum, P. L., Russell, D. J., Walter, S. D., Wood, E. P., et al. (2000). Validation of a model of gross motor function for children with cerebral palsy. *Physical Therapy*, 80, 974–985.
- Pandyan, A. D., Johnson, G. R., Price, C. I., Curless, R. H., Barnes, M. P., & Rodgers, H. A. (1999). Review of the properties and limitations of the Ashworth and modified Ashworth Scales as measures of spasticity. *Clinical Rehabilitation*, 13, 373–383.
- Portney, L. G., & Watkins, M. P. (2000). *Foundations of clinical research* (2nd ed.). Upper Saddle River, New Jersey: Prentice-Hall, Inc.
- R Development Core Team. (2009). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. Vienna, Austria. URL <http://www.R-project.org>.
- Van der Putten, A., Vlaskamp, C., Reyniers, K., & Nakken, H. (2005). Children with profound intellectual and multiple disabilities: The effects of functional movement activities. *Clinical Rehabilitation*, 19, 613–620.
- Shevell, M. D., Dagenais, L., & Hall, N. (2009). Comorbidities in cerebral palsy and their relationship to neurologic subtype and GMFCS level. *Neurology*, 72, 2090–2096.
- Sim, J., & Wright, C. C. (2005). The kappa statistic in reliability studies: Use, interpretation, and sample size requirements. *Physical Therapy*, 85, 257–268.
- Stedman, T. L. (1999). *Stedman's medical dictionary* (27th edition). Philadelphia: Lippincott, Williams and Wilkins Company.
- World Health Organization. (1992). *International statistical classification of diseases and related health problems: Tenth revision (ICD-10)* (pp. 369–370).
- World Health Organization. (2001). *International classification of functioning, disability and health*. WHO: Geneva.
- Yam, W. K., & Leung, M. S. (2006). Interrater reliability of Modified Ashworth Scale and Modified Tardieu Scale in children with spastic cerebral palsy. *Journal of Child Neurology*, 21, 1031–1035.